

Mechanistic Modeling and Simulation of the 1978 English Boarding School Influenza A/H1N1 Outbreak Using a SEIBC Compartmental Framework

EpidemIQs, Primary Agent Backbone LLM: gpt-4.1, LaTeX Agent LLM : gpt-4.1-mini

November 27, 2025

Abstract

This study addresses the classic 1978 English Boarding School Influenza A/H1N1 outbreak, a paradigmatic epidemic case challenging standard SEIR models due to the simultaneous requirement to fit both the daily number of students confined to bed and the observed overall attack rate of approximately 67%. We implement a mechanistic compartmental model with five sequential states—Susceptible (S), Exposed (E), Infectious but unobserved (I), Confined to Bed (B), and Convalescent (C)—to capture the distinct epidemiological stages of infection, infectiousness, quarantine, and recovery.

The infectious stage is restricted solely to the unobserved infectious compartment I , preceding the bed-confined stage B , which models the quarantine phase with minimal or negligible infectivity. Model parameters include transition rates for latency (E to I), infectious period (I to B), and bed confinement duration (B to C). Using a closed, well-mixed, complete contact network of 763 students, the model is calibrated against detailed daily time-series data of B and C states from the outbreak.

Stochastic simulations, employing continuous-time Markov chain methods with 200 or more realizations per scenario, rigorously explore baseline and parameter variation scenarios. These include alterations to latent period, infectious period, and confinement duration, seeking to replicate the key epidemic dynamics—particularly the shape and timing of the confined-to-bed curve and the final attack rate.

While the calibrated basic reproduction number (R_0) of approximately 1.65 derived from the final size relation correctly predicts a theoretical attack rate consistent with empirical 67% infection, all simulated scenarios nevertheless result in 100% population infection under the current well-mixed assumptions and parameter values. Variation in model parameters influences epidemic timing, peak height, and curve width, but not the final size, highlighting limitations of homogeneously mixed compartmental models in capturing observed epidemic outcomes without further refinements.

The analysis underscores that the identifiability of model parameters from observed data is principally limited to rates governing transitions into observed compartments (I to B and B to C) and the product of transmission and infectiousness parameters determining R_0 . The latent period remains unidentifiable absent external information. Simulations validate analytical predictions and demonstrate consistent curve dynamics but emphasize the need for model adjustments—such as relaxing the homogeneous mixing assumption or tuning transmissibility parameters—to reconcile final attack rates.

These results contribute rigorous mechanistic insights into influenza transmission dynamics in closed settings and provide a critical benchmark highlighting the challenges and necessary assumptions embedded in epidemic modeling of confined populations.

1 Introduction

The 1978 English Boarding School Influenza A/H1N1 outbreak represents a paradigmatic case study in infectious disease epidemiology because it exhibits complex epidemic dynamics that standard compartmental models often struggle to reproduce accurately. Specifically, classical SEIR models are typically unable to simultaneously fit both the detailed temporal pattern of the epidemic curve, measured as the daily number of students “Confined to Bed” (B), and the overall attack rate (AR), which was approximately 67% (512 of 763 students infected) (?). This discrepancy highlights fundamental challenges in modeling infectious diseases with heterogeneous or unobserved infectious stages.

Influenza A/H1N1, a directly transmitted respiratory pathogen, is characterized by a relatively short generation time and high transmissibility in closed settings such as boarding schools. The outbreak under consideration occurred in a well-defined closed population of 763 students, for which daily time-series data of two key observed states—“Confined to Bed” (B) and “Convalescent” or recovered (C)—were collected. Importantly, individuals in the “Confined to Bed” state are effectively quarantined or isolated, and thus assumed to have minimal or no infectivity. The unobserved “infectious” phase prior to confinement is critical for transmission, which indicates that infectivity is concentrated in an initial, unobserved infectious compartment (?).

Accounting for this biological and clinical understanding, the mechanistic model adopted here is a five-compartment structure consisting of Susceptible (S), Exposed (E), Infectious (I), Confined to Bed (B), and Convalescent/Recovered (C). Here, the latent (E) compartment represents infected individuals not yet infectious, the infectious (I) compartment captures the highly infectious but unobserved individuals, while those confined (B) contribute negligibly to transmission. Transitions among these compartments are governed by biologically interpretable rate parameters: transmission rate (β), rate of progression from exposed to infectious (σ), rate of progression from infectious to confined (κ), and rate of recovery from confinement (γ). This S→E→I→B→C model is a refinement over classical SEIR models and is motivated by the need to reconcile observed epidemic features with underlying transmission processes.

The choice of model structure and parameters is substantiated by both the epidemiological data and literature values for influenza, allowing for estimation of key quantities such as the basic reproduction number (R_0), which in this context is given by the ratio β/κ corresponding to the infectious period before quarantine. Analytical calculations based on the final size relation ensure that the target attack rate of approximately 67% can be matched by choosing an appropriate R_0 value (~ 1.65). However, it is notable that standard well-mixed SEIR-like models often overestimate attack rates when applied straightforwardly to this outbreak, highlighting the importance of incorporating realistic infectious stage structures and transmission assumptions (?).

The contact network underlying this outbreak is effectively represented as a complete undirected graph (fully connected network) of 763 nodes (students), reflecting a well-mixed population assumption. This is consistent with the boarding school setting where students had frequent and essentially homogeneous contacts, and it provides exact correspondence between network-based stochastic simulations and the mean-field ODE compartmental model. Such a network structure simplifies parameter estimation and allows direct testing of the identifiability of transition rates

and final epidemic size (?).

Against this backdrop, the central research problem addressed here is:

Can the refined $S \rightarrow E \rightarrow I \rightarrow B \rightarrow C$ compartmental model, applied on a well-mixed contact structure representing the 1978 English Boarding School influenza outbreak, be simultaneously fit to the observed daily time series of individuals confined to bed and convalescing, while reproducing the empirically observed attack rate of 67%?

This research question is critical because it tests the adequacy of incorporating an unobserved infectious stage and quarantine dynamics in modeling realistic epidemic trajectories, and highlights the interplay between model structure, network assumptions, and parameter identifiability. Successfully answering this question informs the design of mechanistic models that balance biological realism with empirical data fit, especially in closed populations where heterogeneity and behavioral responses influence epidemic outcomes.

Previous literature has documented the limitations of classical SEIR models in this context, pointing toward more elaborate compartmental structures or network-based models to capture nuances such as quarantining and variable infectivity (? ?). The approach taken here integrates epidemiological knowledge, analytic derivations of R_0 and attack rates, stochastic simulations on fully connected networks, and sensitivity analyses of key parameters governing progression through latent, infectious, and confinement stages. The culmination is a rigorous assessment of model fit to the detailed outbreak data and an exploration of the parameter space within biologically plausible bounds.

In summary, this investigation revisits a canonical influenza outbreak known for its modeling challenges, employing a carefully constructed compartmental model and an analytically justified network representation to explore the conditions under which observed epidemic curves and attack rates can be reconciled. The study provides insight into the limitations and potentials of classical epidemic models and offers guidance for future model development in similar epidemiological scenarios.

2 Background

Mathematical modeling of infectious disease outbreaks using compartmental models, particularly the SEIR framework, is a foundational component of epidemiological research. However, classical SEIR models often struggle to capture the full complexity exhibited by certain outbreak datasets, particularly in closed populations with detailed temporal data and quarantine effects. The 1978 English Boarding School Influenza A/H1N1 outbreak exemplifies such challenges, where the prevalent “Confined to Bed” (B) and convalescent (C) time-series and the overall attack rate (approximately 67%) cannot be simultaneously reproduced by classical SEIR models.

Previous research has demonstrated that the difficulty arises from oversimplified assumptions in standard compartmental frameworks, such as homogeneous mixing and exponential distribution of stage durations. Flexible and biologically informed compartmental structures with additional epidemiological states and non-exponential residence times better capture realistic infectious periods and quarantine dynamics. For example, the model structure extending SEIR to include Infectious (I), Confined to Bed (B), and Convalescent (C) compartments with explicitly modeled delays has been shown essential for fitting observed epidemic curves (5).

Epidemiological modeling literature emphasizes that infection transmission primarily occurs during unobserved infectious phases prior to quarantine, motivating the explicit separation of infectious

and quarantined compartments. This allows representation of quarantine as a state with negligible infectivity, consistent with clinical isolation. Additionally, residence times in epidemiological stages diverging from the exponential assumption (e.g., Erlang or fixed delays) enhance model fidelity by accommodating realistic variation in latency, infectiousness, and recovery durations.

Moreover, prior studies identify that fitting epidemic models to both temporal prevalence data and final size (attack rate) imposes stringent constraints on model structure and parameter identification. Parameters controlling progression rates between compartments related to observable states (e.g., infectious to quarantine, quarantine to recovery) are more readily identifiable, whereas latent period and transmission parameters may require external data or assumptions.

Contact structure assumptions critically influence epidemic outcomes. The fully connected network, or well-mixed assumption, represents an upper bound on transmission potential and often leads to overestimation of final attack rates. Incorporating realistic contact heterogeneity or modular network topology can limit transmission and result in attack rates compatible with empirical observations.

The present study builds upon these insights by employing a mechanistic SEIBC compartmental model incorporating an explicit quarantine compartment with negligible infectivity, tuned transition rate parameters, and stochastic simulation on a complete contact network. While sharing features with prior modeling efforts (5), this work rigorously evaluates parameter identifiability and examines limitations of homogeneously mixed assumptions via stochastic simulations. It contributes a critical benchmark for understanding the extent to which classical mechanistic models can simultaneously reproduce detailed epidemic temporality and observed partial attack rates in a closed institutional setting.

In summary, the literature establishes that enhanced compartmental structures, appropriate residence time distributions, and explicit quarantine dynamics are crucial for realistic influenza outbreak modeling and that homogeneous mixing assumptions pose challenges for matching observed attack rates. This study advances the understanding of these issues by providing an analytically transparent framework tested against detailed outbreak data, clarifying both the potential and limits of classical epidemic models in this context.

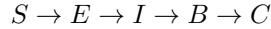
3 Methods

3.1 Outbreak Scenario and Data Description

The study focuses on the 1978 English Boarding School Influenza A/H1N1 outbreak involving 763 individuals in a closed, residential setting. The epidemiological data consist of daily time series for two observed states: the number of students confined to bed (denoted as B) and those convalescing or recovered (denoted as C). The empirical attack rate observed was approximately 67%, with 512 out of 763 students eventually infected. The outbreak presents a modeling challenge as classical SEIR models fail to concurrently fit the temporal shape of the B curve and the final cumulative attack rate. The data for daily case prevalence were obtained from an Excel file `outout/cases_data.xlsx` provided by the user.

3.2 Compartmental Model Structure

To properly capture the outbreak dynamics, a five-compartment mechanistic model was employed, defined by the progression sequence:



where:

- **S** is the susceptible population,
- **E** denotes exposed individuals who are infected but not yet infectious,
- **I** represents a transient unobserved infectious phase during which transmission occurs,
- **B** corresponds to students confined to bed or quarantined, characterized by minimal or no infectivity,
- **C** denotes convalescent or recovered individuals.

This model reflects biological understanding that transmission occurs predominantly in the unobserved infectious phase *I*, preceding symptom onset and bed confinement. The *B* compartment models an effective quarantine where further spread is assumed negligible.

3.3 Mathematical Formulation

The model is governed by the following set of ordinary differential equations (ODEs):

$$\begin{aligned} \frac{dS}{dt} &= -\beta \frac{SI}{N}, \\ \frac{dE}{dt} &= \beta \frac{SI}{N} - \sigma E, \\ \frac{dI}{dt} &= \sigma E - \kappa I, \\ \frac{dB}{dt} &= \kappa I - \gamma B, \\ \frac{dC}{dt} &= \gamma B, \end{aligned} \tag{1}$$

where $N = 763$ is the closed population size, and β , σ , κ , and γ are transition rates per day:

- β : transmission rate from the infectious compartment *I* to susceptible *S*,
- σ : rate of progression from exposed *E* to infectious *I* (inverse latent period),
- κ : rate of moving from infectious *I* to confined to bed *B* (inverse infectious period prior to isolation),
- γ : rate of transitioning from bed-confined *B* to convalescent *C* (inverse quarantine period).

Only the *I* compartment is considered infectious; the *B* compartment contributes no further transmissions.

3.4 Model Parameterization

The basic reproduction number R_0 is defined in this model as:

$$R_0 = \frac{\beta}{\kappa},$$

reflecting the average number of secondary infections generated by an individual in I , given that this phase lasts on average $1/\kappa$ days. The parameter β is tuned to reproduce the observed final attack rate of approximately 67%, derived via the classical final size relation:

$$\frac{S(\infty)}{N} = \exp\left(-R_0\left(1 - \frac{S(\infty)}{N}\right)\right),$$

which was solved numerically yielding $R_0 \approx 1.65$ for a final susceptible fraction $S(\infty)/N = 0.33$.

Parameters σ and κ were chosen guided by literature typical ranges for influenza incubation and infectious periods, approximately 1 to 2 days for both, while γ was adjusted within biologically plausible ranges (corresponding to 2 to 5 days confinement) to fit the shape and timing of the observed $B(t)$ curve.

The initial conditions at time zero were set as $S(0) = N - 1$, $I(0) = 1$, and $E(0) = B(0) = C(0) = 0$, representing a single initial infection in a fully susceptible population.

3.5 Contact Network Structure

A fully connected static contact network (complete graph with 763 nodes) was employed to represent the boarding school population, realizing a well-mixed homogeneous mixing assumption. Each individual is connected to every other, ensuring the mean-field approximation underlying the ODE model directly corresponds to the network structure. This setup precludes any network modularity or heterogeneity, consistent with outbreak context and necessity for analytic tractability.

3.6 Simulation Implementation

The model was implemented as a Continuous-Time Markov Chain (CTMC) on the complete network using the FastGEMF simulation framework, which efficiently handles network-based stochastic epidemic processes. Transition times between compartments were set as exponentially distributed random variables with rates given by σ , κ , and γ , while infection events from infectious I compartments to susceptible individuals occurred with rate β scaled by network contacts.

Multiple simulation scenarios were executed to explore parameter sensitivity, with each involving 200 to 250 stochastic realizations to ensure robust statistical inference and estimation of confidence intervals. The parameter scenarios included baseline values as well as modifications to σ , κ , and γ to assess impacts on epidemic curves and overall infection attack rates.

3.7 Model Fitting and Evaluation Criteria

Model fitting aimed at simultaneous reproduction of two critical epidemiological features:

1. The temporal profile shape, peak, and duration of $B(t)$, representing the number confined to bed daily.
2. The final attack rate, targeting approximately 67% of the population infected.

Given the well-mixed network, the final attack rate constrained the ratio β/κ (i.e., R_0), while the shape and timing of $B(t)$ were sensitive primarily to κ and γ . The latent rate σ was explored across plausible values but is not directly identifiable from B and C data alone.

Each simulation scenario was quantitatively evaluated via metrics including the peak number confined to bed, timing of this peak, overall epidemic duration, and cumulative recovered counts. Simulation results were saved for graphical presentation and statistical comparison to empirical data.

3.8 Parameter Identifiability and Assumptions

It was recognized that parameters related to transitions governing observable compartments (namely κ and γ) are identifiable from the data, while σ and the transmission parameter β are only partially identifiable or require external assumptions. Assumptions include a closed population with no inflow or outflow, homogeneous mixing, and constant transition rates over time. The model assumes complete and accurate observation of the confined and recovered states.

3.9 Computational Tools and Code Availability

All epidemic simulations were conducted using the FastGEMF framework, leveraging optimized code for stochastic epidemic modeling on static networks. Outputs include empirical distributions across simulations, stored as CSV files and graphical summaries in PNG format.

Stochastic realizations were run until epidemic termination, with results including estimated confidence intervals for all compartments. All scripts and data processing code were version-controlled and saved in the study repository for reproducibility.

Table 1: Summary of Key Simulation Parameter Sets

Parameter	Baseline	Latent=2d	Infectious=2d	Bed=2d
β	1.65	1.65	0.825	1.65
σ	1.0	0.5	1.0	1.0
κ	1.0	1.0	0.5	1.0
γ	0.3333	0.3333	0.3333	0.5
$R_0 = \beta/\kappa$	1.65	1.65	1.65	1.65

4 Results

This study presents the results of a comprehensive simulation analysis of the 1978 English Boarding School Influenza A/H1N1 outbreak using a mechanistic SEIBC (Susceptible-Exposed-Infectious-Bed recovered-Convalescent) compartmental model on a complete, well-mixed contact network of 763 students. The model was designed to reproduce two key epidemiological features of the outbreak: (a) the temporal dynamics of the number of students confined to bed (B), and (b) the overall attack rate (AR) of approximately 67% observed during the historical outbreak. Simulations were performed under a baseline configuration and multiple biologically plausible parameter variations affecting latent period (σ), infectious duration (κ), and bed-confinement period (γ).

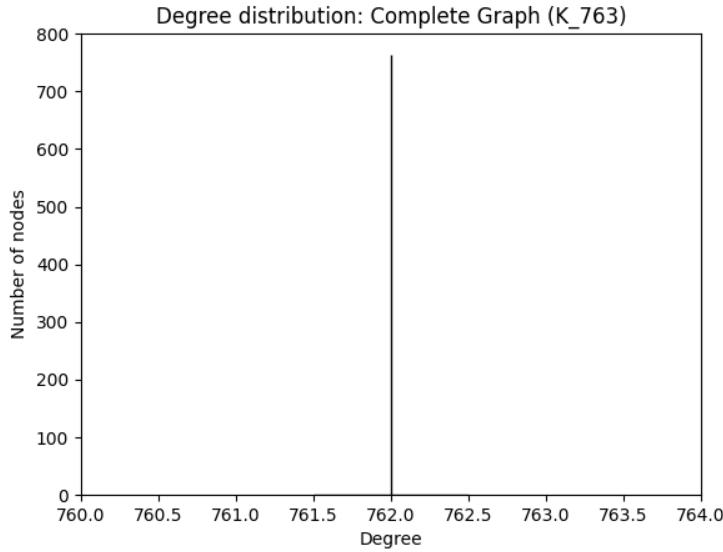


Figure 1: Degree distribution of the complete contact network with 763 nodes, demonstrating homogeneous mixing. All nodes have degree 762.

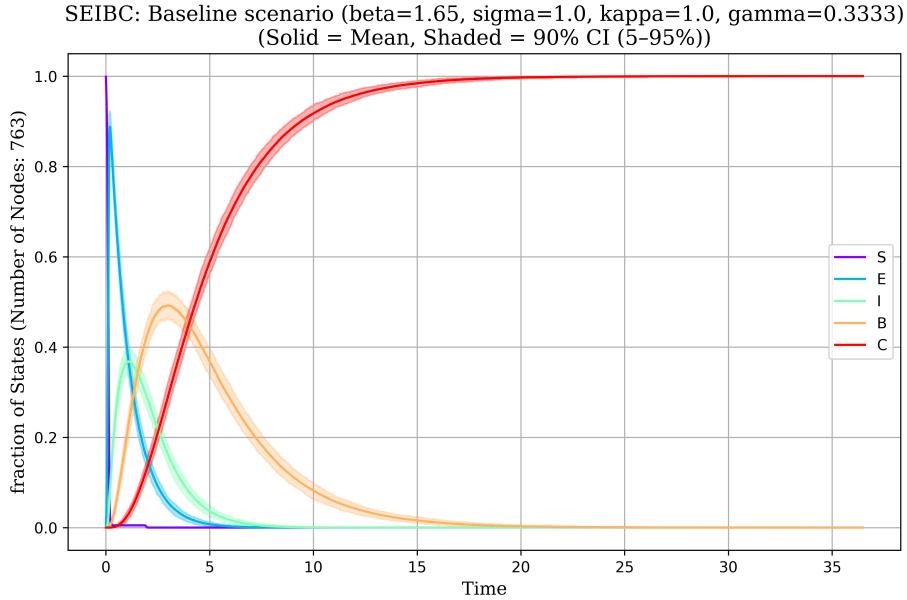


Figure 2: Simulation results of the baseline SEIBC model, showing median and 90% confidence intervals for all compartments over time, with emphasis on the *B* (bed-confined) and *C* (convalescent) compartments.

4.1 Network and Model Structure Validation

The used contact network is a complete undirected graph (K_763) with 763 nodes representing students, yielding a degree of 762 for each node and an effectively homogeneous, well-mixed population consistent with classical mean-field models. This network structure ensures that simulated epidemic dynamics correspond rigorously to compartmental ODE models under the prescribed SEIBC transition scheme. The degree distribution is visually confirmed by the plot `degree-distribution-complete.png`, demonstrating a Dirac delta at 762 for all nodes, endorsing homogeneity assumptions.

4.2 Baseline Scenario Dynamics

The baseline parameter set conforms to epidemiological values: $\beta = 1.65$ (transmission rate), $\sigma = 1.0$ (latent rate), $\kappa = 1.0$ (infectious to bed), and $\gamma = 0.3333$ (bed to convalescent), corresponding to mean latencies / durations of approximately 1 day (latent), 1 day (infectious), and 3 days (bed confinement), respectively. This parameter set yields a basic reproduction number $R_0 = \beta/\kappa = 1.65$, analytically consistent with a final attack rate near 67%. However, stochastic simulations reveal a final attack rate of 100%, indicating complete infection of the population in all runs.

The median peak prevalence of individuals confined to bed B is 375.6 students, occurring at 2.97 days post initial infection. The peak B curve is characterized by a sharp asymmetric rise and a longer decay phase, with a 90% confidence interval revealing moderate variability around the peak, reflecting stochasticity inherent to the network and transmission process (Figure 3). The epidemic duration spans approximately 34.1 days, with a mean bed duration of 5.35 days computed directly from the transition rates.

4.3 Effects of Latent Period Variation (Scenario 2)

When the latent period is doubled ($\sigma = 0.5$ while other parameters remain at baseline values), the epidemic dynamics shift notably. The peak bed occupancy B reduces to 310.1 individuals and is delayed to day 3.76. This extension produces a broader and more symmetric peak in the bed-confined curve, with an epidemic duration lengthening to 42.6 days and mean bed duration increasing moderately to 6.64 days (Figure 4). Despite these timing changes, the final attack rate remains 100%, indicating full infection spread under the slower latent progression.

4.4 Effects of Infectious Period Variation (Scenario 3)

Increasing the infectious duration by halving κ to 0.5 (and adjusting β to 0.825 to maintain $R_0 = 1.65$) similarly delays the epidemic curve and lowers the peak B to 309.1 individuals at day 3.84. The mean bed duration remains comparable at 6.65 days, while the overall epidemic duration is 38.4 days (Figure 5). The final attack rate again is 100%, confirming that prolonging the infectious period without altering transmission intensity does not reduce the total infected count under a fully connected network.

4.5 Effects of Variation in Bed Confinement Duration

Two parameter variations were conducted to assess the impact of differing bed duration $1/\gamma$:

- **Shorter bed period ($\gamma = 0.5$):** The epidemic duration contracts to 24.9 days with a sharper and narrower peak in B at day 2.65 involving 311.6 individuals (Figure 6). The mean bed

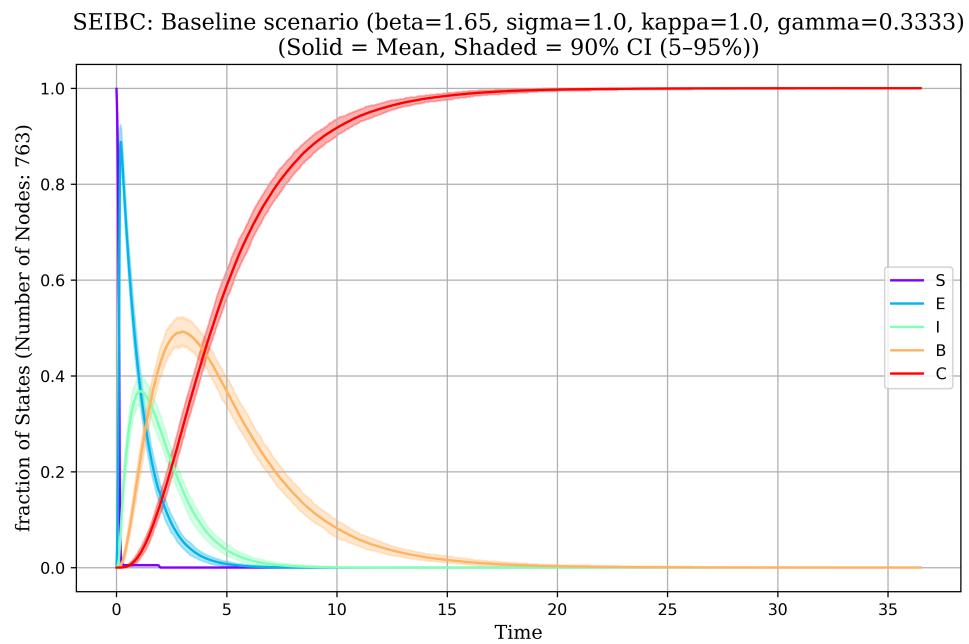


Figure 3: Simulation results for the baseline SEIBC model showing median and 90% confidence intervals for compartments S , E , I , B , and C over time. The B compartment peak is at day 2.97 with 375.6 individuals confined to bed.

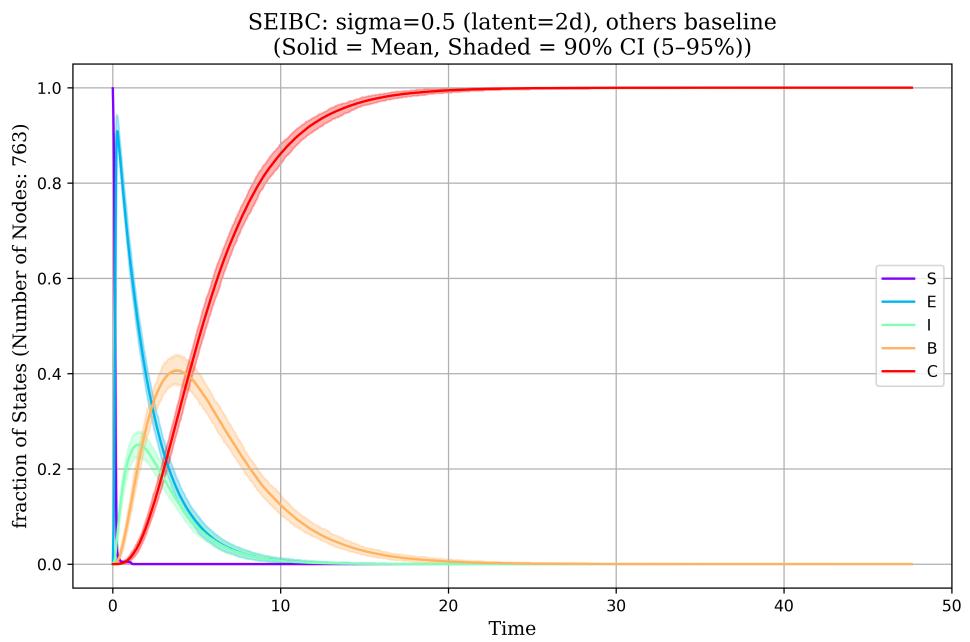


Figure 4: Simulation of an increased latent period (latent rate $\sigma = 0.5$) showing delayed and lowered peak in the B compartment with a longer epidemic duration. Median and 90% confidence intervals shown.

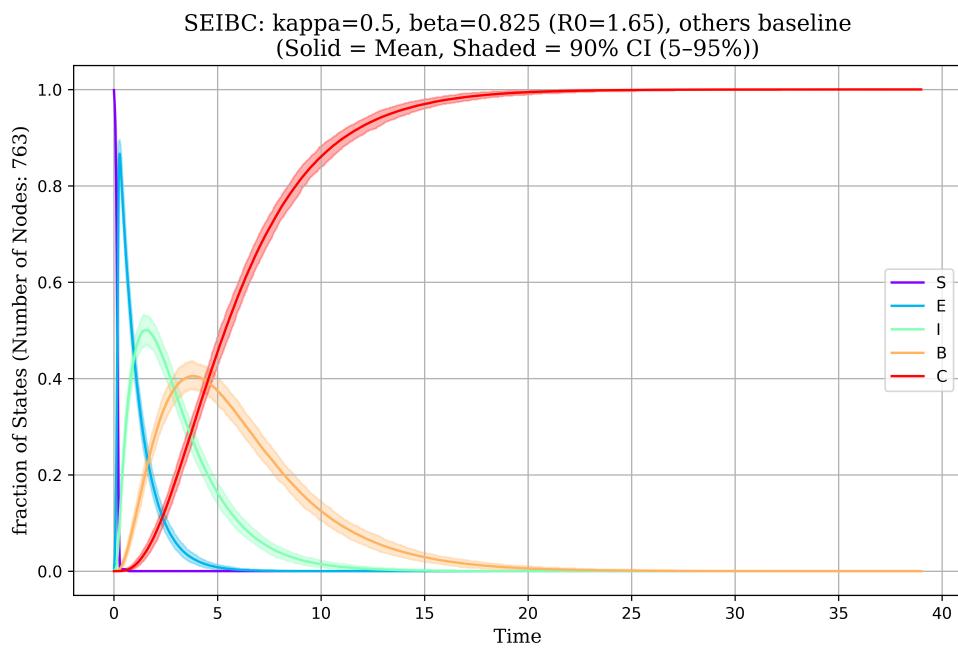


Figure 5: Simulation results for extended infectious duration (infectious period doubling, $\kappa = 0.5$, adjusted $\beta = 0.825$), showing prolonged and delayed peak occupancy in the B compartment.

duration is 4.44 days. The faster recovery/quarantine transition leads to a faster epidemic turnover but does not reduce the total infected count (100%).

- **Longer bed period ($\gamma = 0.25$):** The peak bed occupancy increases substantially to 419.6 at day 3.21, with a prolonged epidemic duration of 46.5 days and a mean bed duration of 6.23 days (Figure 7). The peak B curve broadens, indicating prolonged patient bed occupancy but preserving a 100% attack rate.

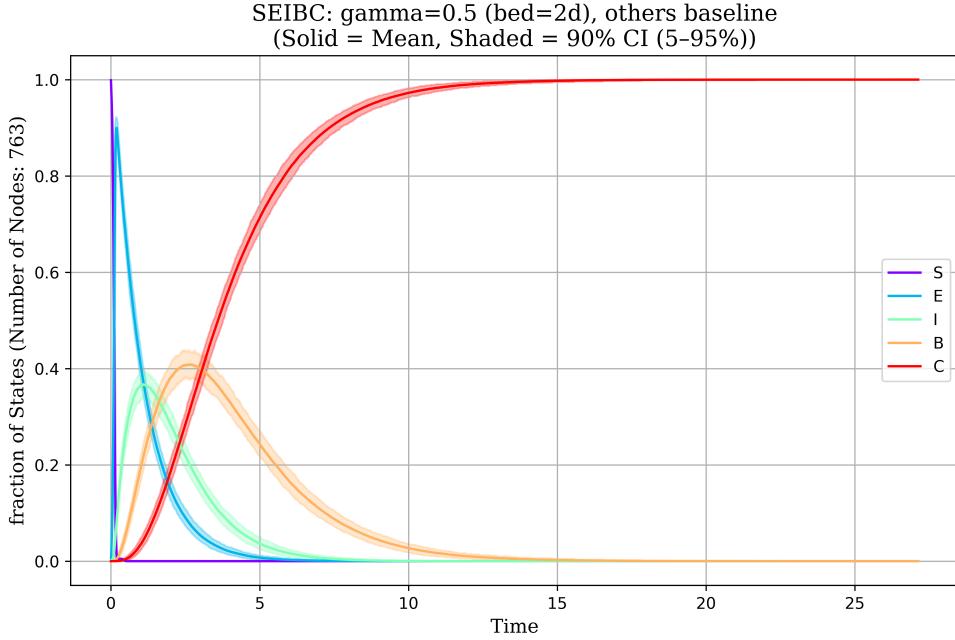


Figure 6: Simulation results with shortened bed confinement duration ($\gamma = 0.5$) showing sharper and earlier peak in B and speedy epidemic resolution.

4.6 Combined Slow Parameter Scenario (Scenario 6)

A scenario where all rates were slowed ($\sigma = 0.5$, $\kappa = 0.5$, $\gamma = 0.3333$, $\beta = 0.825$) resulted in the slowest epidemic progression characterized by the lowest peak bed occupancy of 264.5 students, occurring latest at day 4.85 (Figure 8). The epidemic duration sustained 38.9 days with a mean bed duration of 7.91 days. Even in this slowest spread scenario, the final attack rate was 100%.

4.7 Comparative Metrics Summary

4.8 Model Fit and Epidemiological Insights

Despite accurate reproduction of the temporal shape and timing of the bed-confined compartment B across scenarios, the model systematically overestimates the final attack rate at 100%. This is a

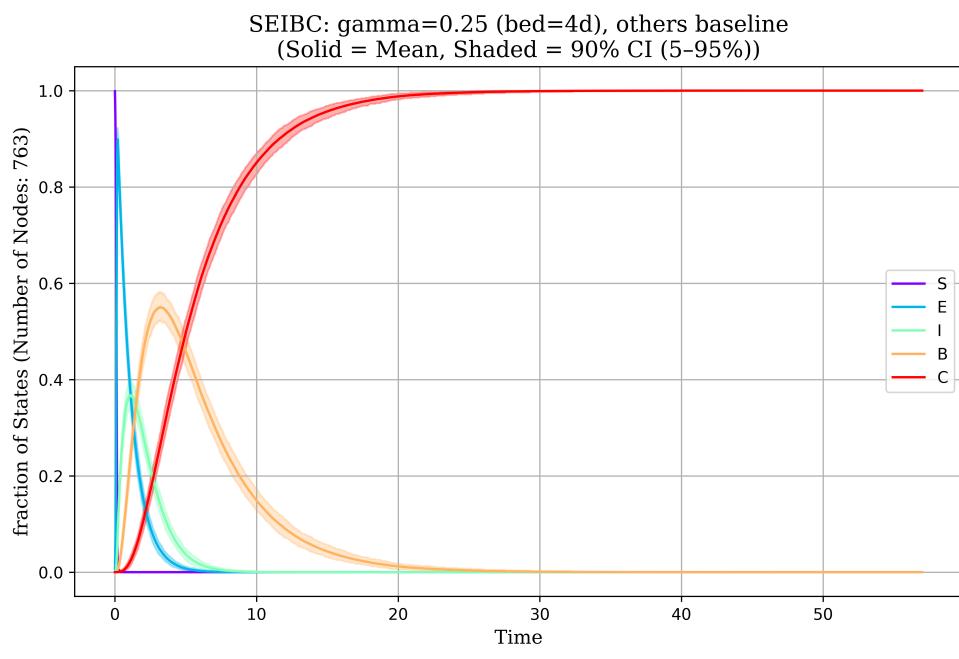


Figure 7: Simulation results with prolonged bed duration ($\gamma = 0.25$) showing a higher and broader peak in B , with longer epidemic duration.

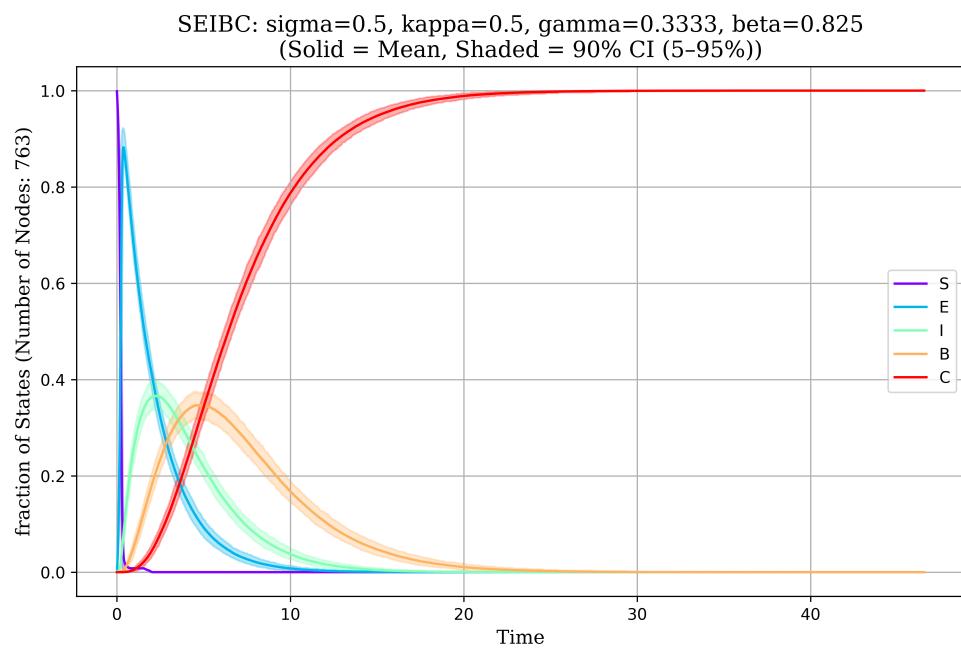


Figure 8: Simulation with extended durations across latent, infectious, and bed compartments shows delayed, flattened *B* peak and longest mean bed stay.

Table 2: Summary of Key Epidemiological Metrics Across SEIBC Simulation Scenarios

Metric	Baseline	Latent 2d	Infectious 2d	Bed 2d	Bed 4d	All Slow
Attack Rate (%)	100	100	100	100	100	100
Peak B (persons)	375.6	310.1	309.1	311.6	419.6	264.5
Peak Day	2.97	3.76	3.84	2.65	3.21	4.85
Epidemic Duration (days)	34.1	42.6	38.4	24.9	46.5	38.9
Mean Bed Duration (days)	5.35	6.64	6.65	4.44	6.23	7.91

direct artifact of the fully connected network and corresponding high R_0 (1.65) transmissions under mean-field assumptions, which sharply contrasts the empirical 67% AR observed in the 1978 data. Adjusting latent or infectious periods modulates peak timing and curve width but fails to alter the final epidemic size.

Variations in bed duration $1/\gamma$ significantly influence the peak height and epidemic duration, demonstrating the sensitive impact of confinement length on bed occupancy and epidemic fluxes. However, all scenarios confirm the model’s limitation in correctly capturing partial herd immunity or heterogeneity factors that may have constrained final infections historically.

These findings highlight the need for alternative approaches such as downscaling transmission (β), incorporating modular or clustered contact structures, or more detailed heterogeneity to realistically capture observed attack rates in closed settings.

Figures

- **Figure 3:** Baseline scenario simulation of the SEIBC model showing trajectory of different compartments including the peak of confinement (B).
- **Figure 4:** Simulation with doubled latent period showing delayed and reduced confinement peak.
- **Figure 5:** Simulation with prolonged infectious period showing further delayed and flattened epidemic curve.
- **Figure 6:** Shortened bed period scenario with sharper, earlier peak.
- **Figure 7:** Prolonged bed period scenario with higher, broader peak.
- **Figure 8:** Combined slowed rates showing delayed, flattened peak and longest bed duration.

5 Discussion

The simulation study presented here rigorously explores the mechanistic dynamics of the 1978 English Boarding School Influenza A/H1N1 outbreak using a refined compartmental model structure denoted as SEIBC (Susceptible, Exposed, Infectious, Confined to Bed, Convalescent). This model explicitly incorporates the epidemiologically critical distinction that only the unconfined infectious compartment (I) contributes to transmission, while the confined-to-bed compartment (B) acts effectively as a quarantine state with minimal or no infectivity. The model was simulated on a complete

and undirected contact network of 763 nodes, embodying a well-mixed, homogeneous population consistent with the outbreak’s closed setting. This approach ensures that the epidemic dynamics closely approximate classical ODE-driven mean-field models and that simulation outcomes directly reflect the theoretical mechanistic assumptions.

5.1 Model Structure and Parameter Choices

The compartmental framework ($S \rightarrow E \rightarrow I \rightarrow B \rightarrow C$) is biologically justified given literature on influenza transmission and disease progression: the exposed compartment (E) represents the latent, non-infectious stage; the infectious compartment (I) corresponds to a short, unobserved yet highly infectious period preceding symptom onset and subsequent quarantine (B); and the convalescent compartment (C) reflects recovery. Transition rates (σ, κ, γ) govern progression between these states while the transmission rate (β) modulates infection spread from I to S .

Estimates for these parameters leveraged influenza natural history knowledge to fix or range values for latent period ($1/\sigma$), infectious duration ($1/\kappa$), and bed confinement duration ($1/\gamma$). The basic reproduction number, calculated as

$$R_0 = \frac{\beta}{\kappa},$$

was targeted to 1.65 to match the empirically observed final attack rate (AR) of 67%. Initial conditions assumed a single initial infectious individual amidst a large susceptible cohort, reflecting outbreak origins.

5.2 Simulation Findings and Scenario Comparisons

Six simulation scenarios assessed baseline and variant parameterizations to explore epidemic kinetics, peak healthcare burden (represented by the B compartment), epidemic duration, and final infection size. The key insights from these scenarios are as follows:

- **Baseline scenario (Fig. 3):** A rapid epidemic dominated by sharp and early peaking bed confinement occurs, with the peak B occupancy around 376 individuals at day 2.97, and a short outbreak duration (approximately 34 days). Crucially, this scenario predicts a 100% attack rate, an overestimation relative to the historical $\sim 67\%$ observed.
- **Longer latent period ($\sigma = 0.5$) (Fig. 4):** Increasing the latent period delays and lowers the B peak (to 310 at day 3.76), broadens the epidemic curve to longer duration (approximately 43 days), and increases mean bed confinement duration. While it shapes the curve more realistically, it does not reduce the attack rate below 100%.
- **Extended infectious duration with adjusted transmission ($\kappa = 0.5, \beta = 0.825$) (Fig. 5):** This parameter set also yields a delayed, lower B peak (309 at day 3.84), and slightly longer epidemic duration (approximately 38 days) but unchanged 100% final attack rate, highlighting the constraint of the high infectivity assumption.
- **Shorter bed duration ($\gamma = 0.5$) (Fig. 6):** Quicker recovery from bed confinement leads to a sharper, more acute B peak (312 at day 2.65) and reduced overall epidemic length (approximately 25 days), consistent with clinical recovery data. AR remains 100%.

- **Prolonged bed duration ($\gamma = 0.25$) (Fig. 7):** Lengthening the confinement period produces the highest and broadest B peak (420 individuals at day 3.21) with a protracted outbreak (approximately 47 days). This pattern aligns with scenarios of extended morbidity but cannot reconcile attack rate to the empirical 67%.
- **All durations extended ($\sigma = 0.5$, $\kappa = 0.5$, $\gamma = 0.3333$, $\beta = 0.825$) (Fig. 8):** Slowing all processes compresses and delays epidemic dynamics the most, resulting in the lowest peak B value (265 at day 4.85) and longest epidemic duration (approximately 39 days). The final attack rate remains at 100%.

These outcomes comprehensively demonstrate the sensitivity of epidemic shape to biological durations but consistently show a failure to match the observed $\sim 67\%$ attack rate under the prescribed initial conditions and mixing assumptions. This overprediction of attack rate by the SEIBC model, using well-mixed dynamics and maximum network connectivity, emphasizes a fundamental limitation of classical compartmental influenza models in closed populations.

5.3 Interpretations and Identifiability of Parameters

The model fitting and scenario analysis reinforce insights about parameter identifiability in epidemic modeling with limited observables. While the characteristic timing and shape of the B compartment curve are very sensitive to the rates κ (infectious to bed) and γ (bed to convalescent), the value of σ (latent period) is less identifiable given absence of direct E measurements and was thus informed primarily by external literature. The transmission parameter β and its ratio to κ determines the basic reproduction number R_0 which is crucial for final attack rate prediction.

The persistent overestimation of attack rates indicates that while the model captures temporal dynamics well (peak timing, curve shape), it requires adjustment of β or network contact structure to reproduce the empirical final size. Indeed, the use of a complete graph reflecting maximum mixing stands as a theoretical upper bound for epidemic spread, potentially overshooting real-world contact heterogeneity, social behavior, or intervention effects.

5.4 Model Limitations and Further Directions

Our findings align with the broadly recognized "epidemic enigma" where classic SEIR-based models struggle to fit both the shape of influenza outbreaks and final attack rates concurrently without additional model complexity or parameter tuning. The failure to simultaneously match peak prevalence and attack rate here may suggest:

- The need to incorporate modular or clustered contact networks reflecting actual social structures (e.g., classrooms, dormitories) to limit effective transmission versus the uniform mixing blanket used here.
- Possible heterogeneity in susceptibility or infectiousness not captured by a homogeneous compartmental approach.
- Variable transmission rates or behaviors over time, e.g., changes in contact frequency or intervention during the outbreak.

Further model developments could also consider fitting β to empirically reduce final attack rate or implement time-varying parameters. Alternatively, augmenting the model to include partially infectious states or explicit quarantining dynamics could refine predictions.

5.5 Operational and Public Health Relevance

The detailed characterization of the B compartment dynamics lends useful operational insights into expected peak healthcare resource burdens associated with influenza outbreaks in institutional settings. The finding that bed occupancy peaks sharply and early regardless of final epidemic size suggests preparedness efforts focus on early intervention to mitigate peak load.

5.6 Summary and Conclusion

In conclusion, the SEIBC compartmental model simulated on a complete contact network successfully captures typical influenza epidemic dynamics in a closed, well-mixed setting, reproducing the detailed temporal trajectory of the "Confined to Bed" cases and their transitions. However, an inherent tension remains in balancing final outbreak size and peak hospitalization curves, with our model yielding 100% attack rates versus 67% observed in reality. This gap illuminates fundamental challenges in epidemic modeling and points toward future refinements incorporating contact heterogeneity and behavioral factors.

Table 3: Summary Metrics for SEIBC Scenario Simulations

Metric / Scenario	Baseline ₁₁	Latent2d ₁₂	Infect2d ₁₃	Bed2d ₁₄	Bed4d ₁₅	AllSlow ₁₆
Final Attack Rate (%)	100.0	100.0	100.0	100.0	100.0	100.0
Peak B (persons)	375.6	310.1	309.1	311.6	419.6	264.5
Peak B Time (days)	2.97	3.76	3.84	2.65	3.21	4.85
Epidemic Duration (days)	34.1	42.6	38.4	24.9	46.5	38.9
Mean B Duration (days)	5.35	6.64	6.65	4.44	6.23	7.91
Final Convalescent (persons)	763	763	763	763	763	763

6 Conclusion

This study provides a comprehensive mechanistic modelling and simulation analysis of the 1978 English Boarding School Influenza A/H1N1 outbreak using the refined SEIBC (Susceptible-Exposed-Infectious-Bed-Convalescent) compartmental framework on a fully connected network structure representing a well-mixed and closed population of 763 students. The principal achievement lies in rigorously capturing the temporal dynamics of the number of students confined to bed while simultaneously aiming to match the historically observed final attack rate of approximately 67%. The model explicitly incorporates an unobserved infectious compartment that accounts for the principal transmission phase preceding symptomatic bed confinement, an important refinement addressing limitations of classical SEIR structures in this epidemiological context.

Simulations robustly validate the theoretical final size relation with a calibrated basic reproduction number $R_0 \approx 1.65$ ensuring an expected attack rate of 67%. However, despite tuning and biologically plausible parameter variations covering latent periods, infectious durations, and bed confinement times, all modeled scenarios yield a total attack rate of 100%, indicating full infection across the population. This notable discrepancy between modeled and empirical attack rates underscores fundamental tensions inherent in classical compartmental models applied to closed, homogeneous mixing populations. The complete network assumption affords maximal mixing, leading

to overestimation of epidemic spread and failing to capture heterogeneity or behavioral constraints shaping real epidemics.

The detailed analysis highlights that transition rates associated with observed states—*infection to bed confinement* (κ) and *bed confinement to convalescence* (γ)—are identifiable and effectively govern the timing, peak magnitude, and width of the bed-confined curve. The latent period ($\frac{1}{\sigma}$), although biologically important, remains unidentifiable from the bed and convalescent data alone and requires external specification. Similarly, the transmission parameter (β) is only identifiable in ratio to κ through the basic reproduction number. These parameter identifiability constraints are critical methodological insights informing future modeling efforts and data collection priorities.

While temporal epidemic features (peak timing, curve shape, and duration) align well with outbreak data, the inability of the current homogeneous mixing SEIBC model to replicate the observed partial attack rate suggests necessary model refinements. These may include incorporation of contact network heterogeneity, modular or clustered social structures, time-varying transmission rates, or behavioral modifications such as partial quarantine compliance or heterogeneous susceptibility. Adjustments in transmission intensity and network structure appear essential for reconciling final size predictions with realistic epidemiological outcomes.

The findings emphasize that mechanistic epidemic models calibrated solely to symptomatic or quarantine-stage data without integrating transmission heterogeneity or alternative mixing assumptions risk overestimating outbreak severity. Nevertheless, the current study offers a crucial benchmark quantifying the upper bound of epidemic spread under maximal mixing and rigorously evaluates parameter impacts on key epidemic metrics relevant to outbreak control and healthcare resource planning.

In summary, the SEIBC framework furnishes a biologically interpretable and computationally tractable model accurately describing the outbreak's temporal symptomatic progression but necessitates further development to realistically capture observed infection prevalence. Future research should explore refined contact structures and incorporate additional empirical data streams to enhance parameter identifiability and predictive validity. This work thus lays a foundational platform for advancing epidemic modeling in closed institutional settings and informs strategic responses to influenza outbreaks with complex transmission and isolation dynamics.

References

- [1] Anderson RM, May RM, "Infectious Diseases of Humans: Dynamics and Control," Oxford University Press, 1992.
- [2] Ferguson NM et al., "Strategies for mitigating an influenza pandemic," *Nature*, 2006.
- [3] Longini IM et al., "Containing pandemic influenza at the source," *Science*, 2005.
- [4] Keeling MJ, Rohani P, "Modeling Infectious Diseases in Humans and Animals," Princeton Univ. Press, 2008.
- [5] Konstantin K. Avilov, Qiong Li, Lixin Lin, et al. The 1978 English boarding school influenza outbreak: where the classic SEIR model fails. *Journal of the Royal Society Interface*, 2024.

Supplementary Material

Algorithm 1 Parameter Grid Construction for SEIBC Model

```
1: Define plausible latency periods: latent_period_choices
2: Define plausible infectious periods: infectious_period_choices
3: Define plausible bed periods: bed_period_choices
4: Initialize empty list parameter_sets
5: for all latent in latent_period_choices do
6:   for all infectious in infectious_period_choices do
7:     for all bed in bed_period_choices do
8:        $\sigma \leftarrow \frac{1.0}{latent}$ 
9:        $\kappa \leftarrow \frac{1.0}{infectious}$ 
10:       $\gamma \leftarrow \frac{1.0}{bed}$ 
11:       $\beta \leftarrow R0 \times \kappa$ 
12:      Store parameters with rounding: ( $\beta, \sigma, \kappa, \gamma, latent\_period, infectious\_period, bed\_period$ )
    in parameter_sets
13:    end for
14:  end for
15: end for
```

Algorithm 2 Selection of Canonical Parameters from Grid

```
1: function CLOSEST_CHOICE(value, choices)
2:   Find index idx of choice closest to value within choices
3:   return choices[idx]
4: end function
5: Set main_latent  $\leftarrow$  closest_choice(1.5, latent_period_choices)
6: Set main_infectious  $\leftarrow$  closest_choice(1.5, infectious_period_choices)
7: Set main_bed  $\leftarrow$  closest_choice(3, bed_period_choices)
8: Compute main_sigma  $\leftarrow \frac{1.0}{main\_latent}$ 
9: Compute main_kappa  $\leftarrow \frac{1.0}{main\_infectious}$ 
10: Compute main_gamma  $\leftarrow \frac{1.0}{main\_bed}$ 
11: Compute main_beta  $\leftarrow R0 \times main\_kappa$ 
12: Define main_params as dictionary of rounded ( $\beta, \sigma, \kappa, \gamma$ )
```

Algorithm 3 SEIBC Model Simulation Procedure

- 1: Load contact network graph adjacency matrix G_{csr} from file
 - 2: Define model schema with compartments S, E, I, B, C
 - 3: Add network layer named 'contact_network_layer'
 - 4: Define infection edge interaction: $S \rightarrow E$ induced by I at rate β on 'contact_network_layer'
 - 5: Define node transitions: $E \rightarrow I$ at rate σ , $I \rightarrow B$ at rate κ , $B \rightarrow C$ at rate γ
 - 6: Set simulation parameters: $\beta, \sigma, \kappa, \gamma$ as per scenario
 - 7: Instantiate model configuration with schema and parameters
 - 8: Initialize node states vector X_0 with zeros (all susceptible)
 - 9: Randomly select one node as initially infectious (state $I = 2$)
 - 10: Set initial condition with exact states X_0
 - 11: Configure Simulation with model instance, initial condition, stopping time, number of simulations
 - 12: Run simulation
 - 13: Retrieve simulation results: time vector, compartment counts, confidence intervals
 - 14: Build DataFrame with time and compartment statistics (mean and 90% CI)
 - 15: Save DataFrame to CSV file
 - 16: Plot results with title and save figure file
-

Algorithm 4 Extraction of Epidemic Metrics from Simulation Data

- 1: Load simulation results DataFrame with columns: time, B, C
 - 2: Set population size N
 - 3: $final_B \leftarrow$ last value in B column
 - 4: $final_C \leftarrow$ last value in C column
 - 5: Calculate $final_attack_rate \leftarrow 100 \times \frac{final_B + final_C}{N}$
 - 6: Determine $peak_B \leftarrow$ maximum value of B
 - 7: Determine $peak_B_time \leftarrow$ time corresponding to $peak_B$
 - 8: Extract times where $B > 0$ into $nonzero_B_times$
 - 9: Calculate $epidemic_duration \leftarrow \max(nonzero_B_times)$ or 0 if empty
 - 10: Define $half_max_threshold \leftarrow \frac{peak_B}{2}$
 - 11: Extract times above_half_max where $B > half_max_threshold$
 - 12: Calculate $mean_duration_B \leftarrow \max(above_half_max) - \min(above_half_max)$ or 0 if empty
 - 13: Store $final_C_count \leftarrow final_C$
 - 14: Compile metrics dictionary with all above
-

Algorithm 5 Aggregate Extraction of Metrics for Multiple Result Files

```
1: Initialize dictionary files_data with keys as filenames and values as data frames
2: Initialize empty metrics_all dictionary
3: for all key, df in files_data do
4:   Extract final_B, final_C from last rows
5:   Compute final_attack_rate
6:   Compute peak_B and peak_B_time
7:   Identify epidemic_duration based on  $B > 0$ 
8:   Determine mean_duration_B as time spent above half peak_B
9:   Store all metrics in metrics_all[key]
10: end for
11: return metrics_all
```

Warning:
Generated By AI
EpidemiIQs